Machine Learning 2 - Homework Assignment 4

Alexandra Lindt

October 2, 2019

Problem 1

1.

$$p(Z, X) = p(z_1)p(x_1|z_1)\prod_{i=2}^{N} p(z_i|z_{i-1})p(x_i|z_i)$$

2.

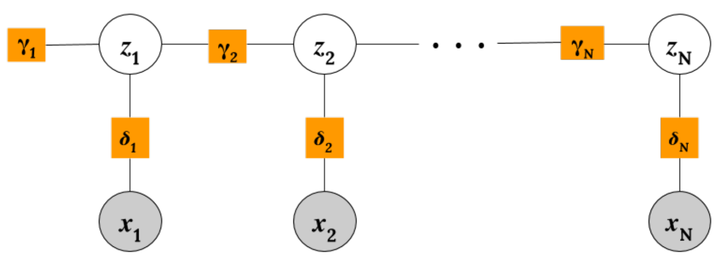

Figure 1: Corresponding factor graph; $\gamma_i, \delta_i \;\; \forall i \in \{1, \dots, N\}$ denote the factors.

3.

$$p(Z, X) = f_{\gamma_1}(z_1)f_{\delta_1}(x_1, z_1)\prod_{i=2}^{N} f_{\gamma_i}(z_i, z_{i-1})f_{\delta_i}(x_i, z_i)$$

4.

$$
\begin{aligned}
p(z_n|X) &= \frac{p(X|z_n)p(z_n)}{p(X)} = \frac{p(x_1,\dots,x_N|z_n)p(z_n)}{p(X)} \overset{\star}{=} \frac{p(x_1,\dots,x_n|z_n)p(x_{n+1},\dots,x_N|z_n)p(z_n)}{p(X)} \\
&= \frac{p(x_1,\dots,x_n|z_n)p(x_{n+1},\dots,x_N,z_n)}{p(X)} = \frac{\alpha(z_n)\beta(z_n)}{p(X)}
\end{aligned}
$$

$\star$: An observed z_n d-separates the graph and makes the variables $\{x_1,\dots,x_n\}$ independent from the variables $\{x_{n+1},\dots,x_N\}$.

$$
\begin{aligned}
\alpha(z_n) &= p(x_1,\dots,x_n|z_n) \\
&= \frac{p(x_1,\dots,x_n,z_n)}{p(z_n)} \\
&= \sum_{z_{n-1}} \frac{p(x_1,\dots,x_n,z_n|z_{n-1})p(z_{n-1})}{p(z_n)} \\
\star_1 &= \sum_{z_{n-1}} \frac{p(x_1,\dots,x_{n-1}|z_{n-1})p(x_n,z_n|z_{n-1})p(z_{n-1})}{p(z_n)} \\
&= \sum_{z_{n-1}} p(x_1,\dots,x_{n-1}|z_{n-1})p(x_n,z_{n-1}|z_n) \\
&= \sum_{z_{n-1}} p(x_n,z_{n-1}|z_n)\alpha(z_{n-1})
\end{aligned}
$$

$\star_1$: An observed z_{n-1} d-separates $\{x_1,\dots,x_{n-1}\}$ from $\{x_n,z_N\}$.

$$
\begin{aligned}
\beta(z_n) &= p(x_{n+1},\dots,x_N,z_n) \\
&= \sum_{z_{n+1}} p(x_{n+1},\dots,x_N,z_n|z_{n+1})p(z_{n+1}) \\
\star_2 &= \sum_{z_{n+1}} p(x_{n+1},z_n|z_{n+1})p(x_{n+2},\dots,x_N|z_{n+1})p(z_{n+1}) \\
&= \sum_{z_{n+1}} p(x_{n+1},z_n|z_{n+1})p(x_{n+2},\dots,x_N,z_{n+1}) \\
&= \sum_{z_{n+1}} p(x_{n+1},z_n|z_{n+1})\beta(z_{n+1})
\end{aligned}
$$

$\star_2$: An observed z_{n+1} d-separates $\{x_{n+1},z_N\}$ from $\{x_{n+2},\dots,x_N\}$.

Problem 2

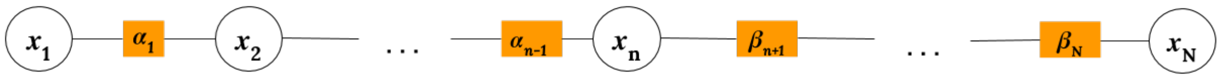

Figure 2: The chain of nodes model considered in this problem.

1.

Starting inference from the left side, we can write :

$$
\begin{aligned}
&\mu_{1\to\alpha_1}(x_1) = 1\\
&\mu_{\alpha_1\to 2}(x_2) = \sum_{x_1}\psi_{1,2}(x_1,x_2)\mu_{1\to\alpha_1}(x_1) = \sum_{x_1}\psi_{1,2}(x_1,x_2)\\
&\mu_{2\to\alpha_2}(x_2) = \mu_{\alpha_1\to 2}(x_2)\\
&\mu_{\alpha_2\to 3}(x_3) = \sum_{x_2}\psi_{2,3}(x_2,x_3)\mu_{2\to\alpha_2}(x_2)\\
&\qquad\qquad\vdots\\
\implies\quad &\mu_{i\to\alpha_i}(x_i) = \begin{cases} 1 & \text{if}\quad i=1\\ \mu_{\alpha_{i-1}\to\ i}(x_i) & \forall i\in\{2,\dots,n\}\end{cases}\\
&\mu_{\alpha_{i-1}\to i}(x_i) = \sum_{x_{i-1}}\psi_{i-1,i}(x_{i-1},x_i)\mu_{i-1\to\ \alpha_{i-1}}(x_{i-1}) \qquad \forall i\in\{2,\dots,n\}
\end{aligned}
$$

In analogy, if we're starting inference from the right side:

$$
\begin{aligned}
&\mu_{N\to\beta_N}(x_N) = 1\\
&\mu_{\beta_N\to N-1}(x_{N-1}) = \sum_{x_N}\psi_{N-1,N}(x_{N-1},x_N)\mu_{N\to\beta_1}(x_N) = \sum_{x_N}\psi_{N-1,N}(x_{N-1},x_N)\\
&\mu_{N-1\to\beta_{N-1}}(x_{N-1}) = \mu_{\beta_N\to N-1}(x_{N-1})\\
&\qquad\qquad\vdots\\
\implies\quad &\mu_{i\to\beta_i}(x_i) = \begin{cases} 1 & \text{if}\quad i=N\\ \mu_{\beta_{i+1}\to\ i}(x_i) & \forall i\in\{n,\dots,N-1\}\end{cases}\\
&\mu_{\beta_{i+1}\to i}(x_i) = \sum_{x_{i+1}}\psi_{i,i+1}(x_i,x_{i+1})\mu_{i+1\to\ \beta_{i+1}}(x_i) \qquad \forall i\in\{n,\dots,N-1\}
\end{aligned}
$$

By putting both equations together, we can write the marginal of x_n using normalizing constant Z as

$$p(x_n) = \frac{1}{Z}\mu_{\alpha_{n-1}\to n}(x_n)\mu_{\beta_{n+1}\to n}(x_n) = \frac{1}{Z}\mu_{\alpha}(x_n)\mu_{\beta}(x_n)$$

Problem 4

Using Bishop formulas (8.61), (8.63) and (8.69) and definitions from chapter 8.4.4:

$$\begin{aligned}
&f_s(x_s) \prod_{i\in\mathrm{ne}(f_s)} \mu_{x_i\to f_s}(x_i) \\
&= f_s(x_s) \prod_{i\in\mathrm{ne}(f_s)} \prod_{l\in\mathrm{ne}(x_i)\setminus f_s} \mu_{f_l\to x_i}(x_i) \\
&= f_s(x_s) \prod_{i\in\mathrm{ne}(f_s)} \prod_{l\in\mathrm{ne}(x_i)\setminus f_s} \sum_{X_{il}} F_l(x_i, X_{il}) \\
&= \sum_{x\setminus x_s} f_s(x_s) \prod_{i\in\mathrm{ne}(f_s)} \prod_{l\in\mathrm{ne}(x_i)\setminus f_s} F_l(x_i, X_{il}) \\
&= \sum_{x\setminus x_s} p(x) \\
&= p(x_s)
\end{aligned}$$